SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14734
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Strategic Clustering of Poverty Areas in Central Java Using K-Means and Silhouette Evaluation

Chusnuut Tacharri[1)], Asih Rohmani[2], Amiq Fahmi[3)] *
[1,2,3)] Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
[1)] 112202106620@mhs.dinus.ac.id, [2)] aseharsoyo@dsn.dinus.ac.id, [3)] *amiq.fahmi@dsn.dinus.ac.id

Submitted :Apr 30, 2025 | **Accepted** : May 18, 2025 | **Published** : May 22, 2025

Abstract: Indonesia is one of several developing nations that struggle with poverty. Central Java is one of Indonesia's provinces with the third-highest percentage of the country's inadequate. This study aims to explore and improve the use of the K-Means Algorithm to study socioeconomic differences. In this study, the Elbow technique is used to estimate the ideal number of clusters to address the shortcomings in estimating the number of clusters in standard K-Means. Model evaluation using the silhouette coefficient shows the effectiveness of this method approach with a value of 0.504 and several clusters (K = 3), which meets the medium structure category. The Human Development Index (HDI) and Uninhabitable Households (RTLH) were two criteria used to categorize poverty areas using the K-Means Algorithm optimization successfully. According to the clustering results, there were 12 regions in Cluster 0, 2 in Cluster 1, and 21 in Cluster 2. These findings are anticipated to offer the Central Java Provincial Government critical insights, facilitating the development of precise and well-targeted initiatives to address deprivation issues effectively. Moreover, a more systematic and structured optimization of the K-Means technique could substantially enhance both the accuracy and practical significance of research on socioeconomic inequality in Central Java Province. This enhanced methodological approach can provide more in-depth results on data-driven regional disparities to reduce these disparities comprehensively.

Keywords: Central Java; Elbow Method; K-Means Clustering; Poverty; Silhouette Coefficient

INTRODUCTION

Poverty is a global phenomenon that requires extraordinary efforts to overcome, becoming a critical issue in every country, especially in developing countries such as Indonesia (Qulsum et al., 2024;Wirayudha & Prasetyia, 2024). Based on data from Badan Pusat Statistik (BPS), the Indonesian Central Statistics Agency, the number of individuals experiencing poverty in Indonesia in March 2023 was recorded at 25.90 million, down from 26.36 million in March 2022. Despite the decline, the percentage of poor people remains relatively high, at 9.36% (Purwanti, 2024). One of the provinces significantly affected is Central Java Province, which is ranked third with the biggest number of poor people in Indonesia. The poverty rate in urban and rural areas of Central Java reached 10.93% in 2023. This percentage is still considerably very high and requires further attention Through examining patterns of poverty distribution, there remain significant opportunities for research to contribute to the formulation of policies that are more impactful and resource-efficient in addressing poverty issues in Central Java.

The intricate socio-economic variations in Central Java emphasize the critical importance of employing clustering methods to gain deeper insights into regional disparities and potential. As an effective strategy for poverty alleviation, clustering analysis, a cultivated data mining technique, uncovers patterns and generates actionable insights, enabling evidence-based decision-making and facilitating more precise, targeted, and impactful interventions (Lasfeto et al., 2024). This clustering method groups regions with similar poverty levels based on variables influencing poverty (Sepriyanti et al., 2022) and categorizes objects sharing common characteristics into cohesive groups (Lubis & Hendrik, 2023). The K-Means algorithm, widely used in data mining, facilitates the grouping of areas in Indonesia, including Central Java Province, according to indicators such as the Human Development Index (HDI) and Uninhabitable Houses (RTLH), supporting the formulation of precise and impactful poverty alleviation policies (Azzahro et al., 2024).

The K-Means algorithm is a partition-based clustering technique that allocates data points to clusters according to their shortest distance to the nearest centroid. The clustering procedure involves calculating the

*Amiq Fahmi

centroid for each cluster and allocating data points according to their proximity to these centroids (Dubey & Rajavat, 2023). One of the main drawbacks of this algorithm is the difficulty in identifying the most appropriate number of clusters (K), as an inaccurate selection can result in less effective clustering outcomes. This limitation can be mitigated using the Elbow Method, which estimates the optimal cluster count by examining the proportion of variance explained as the total number of clusters expands. The generated plot displays a noticeable inflection point resembling an elbow, which signifies the optimal number of clusters (Onumanyi et al., 2022). This method offers a structured and dependable means of identifying the optimal cluster count, thereby strengthening the validity of the clustering analysis.

The objective of this study is to perform a socio-economic disparity analysis by clustering cities and regencies in Central Java Province based on poverty indicators using an optimized K-Means algorithm. Previous studies have largely overlooked the incorporation of clustering evaluation models utilizing accuracy metrics such as the Silhouette Coefficient and the comprehensive integration of socio-economic indicators within the analytical process. We improve the Elbow Method to more accurately identify the ideal cluster count, addressing the limitations of the traditional K-Means algorithm. Furthermore, the outcome of the clustering process are validated using the Silhouette Coefficient to evaluate their quality and accuracy. The integration of the Elbow Method and Silhouette Coefficient provides a more precise evaluation than conventional approaches, enhancing the significance of this study in delivering highly accurate clustering results. As an added point, the approach employed in this study, which integrates the Elbow Method and Silhouette Coefficient, has been scarcely applied in poverty analysis, offering an innovative contribution within this context. Ultimately, this study can provide valuable insights for the Central Java Provincial Government to formulate more focused and targeted poverty alleviation policies. In a more comprehensive framework, using a systematic elbow and silhouette coefficient method to improve the K-Means algorithm, this study can make the socio-economic disparity clustering analysis in Central Java Province more accurate and useful.

LITERATURE REVIEW

Poverty, defined as the inability to meet basic needs, is a complex problem to overcome and is often caused by economic incapacity, which hampers national and regional growth in Indonesia (Maesarah & Fatah, 2024), (Arinto Umbu Dasa et al., 2024). Clustering algorithms have been widely used to analyze poverty data in Indonesia. These studies aim to support the decision-making process for poverty alleviation programs by grouping provinces or households based on poverty indicators. K-means consistently outperforms other clustering methods in poverty analysis, such as DBSCAN and K-Medoids (Vanessa et al., 2024; F. Andriyani & Puspitarani, 2022).The optimal number of clusters varies across studies, ranging from three to five, depending on the specific context and data used (W. Andriyani et al., 2023; Riyono & Pujiastuti, 2022). This clustering analysis has proven helpful in identifying poverty patterns, grouping areas based on poverty levels, and informing targeted poverty alleviation policies and programs. Applying data mining techniques to poverty data offers a more efficient and in-depth understanding of poverty distribution in Indonesia.

Maesarah and Fatah used the K-means clustering algorithm to group Indonesian regions into several clusters of poverty-affected areas based on poverty levels (Maesarah & Fatah, 2024). The data used came from Kaggle, including the quantity of provinces in Indonesia and the quantity of poverty-affected populations in each province. They used the performance operator to determine the optimal cluster. The Davis-Bouldin Index (DBI) evaluated the four poverty clusters this study produced in Indonesia. Dasa et al. (Arinto Umbu Dasa et al., 2024) implemented the K-Means clustering algorithm to categorize poverty-affected populations by village in Tana Righu District, aiming to mitigate the inefficiencies in aid distribution and ensure more targeted assistance. Data was collected from poverty-related population data in Tana Righu sub-district villages, and the optimal cluster was determined using the Davis-Bouldin Index (DBI). This study produced six clusters divided based on poverty levels according to specific characteristics, with the evaluation of clustering results using the Cluster Distance Performance operator. However, these studies have not discussed the results of evaluating the resulting clusters.

The K-Means algorithm was used by Anggraeni and Enri (R et al., 2022) and Bahauddin et al. (Bahauddin et al., 2021) to group poverty data in Indonesian provinces, especially in West Java, based on the number of poverty-affected populations, the Human Development Index (HDI), and the open unemployment rate (Amelia et al., 2023). The data collected were secondary data from the BPS website of West Java Province. In this study, clusters were determined using the elbow approach, which involves assessing the variation within each cluster and choosing the point when adding another cluster does not significantly enhance the model. This study produced 3 clusters, with each cluster having different characteristics. Specifically, cluster 1 was included in the reasonably high poverty group, cluster 2 in the low poverty group, and cluster 3 in the high poverty group. This study also did not discuss the results of the cluster evaluation.

Previous studies have widely carried out the K-Means algorithm to cluster poverty-affected areas, demonstrating its effectiveness in identifying spatial socio-economic disparities. However, few studies discuss evaluating clusters with the silhouette coefficient. This evaluation aims to determine the accuracy level of the cluster results that have been produced. In this context, our study tries to fill the gap by clustering poverty-affected

*Amiq Fahmi

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i2.14734 p-ISSN : 2541-044X

areas based on the Human Development Index (HDI) and the quantity of Uninhabitable Houses (RTLH), giving a more comprehensive understanding of socioeconomic disparities. It utilizes the K-Means algorithm and the Elbow Method, the best way to find the best groups. This method enables our study to compare multiple clusters based on the Sum of Squared Errors (SSE), facilitating a more robust assessment of clustering accuracy and performance. Furthermore, analysis and evaluation of the resulting clusters are analyzed using the Silhouette Coefficient to determine the cluster results' accuracy level.

METHOD

This study consists of five primary steps: data collecting, data preparation, use of the Elbow method, K-means clustering, and evaluation. Fig 1 details each stage.

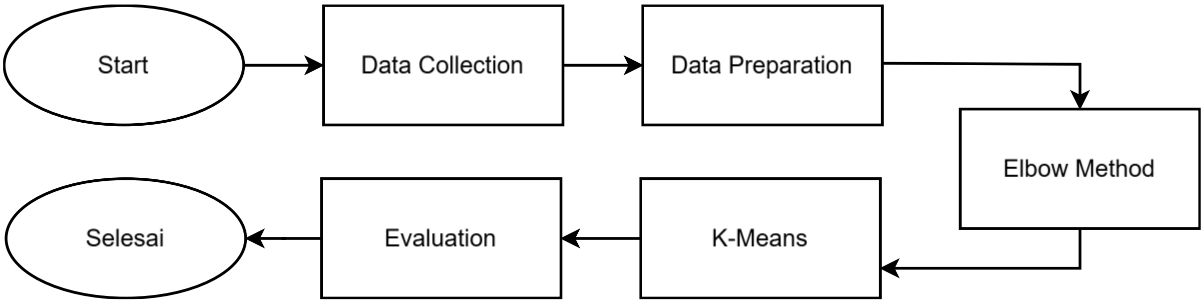

Fig 1. Methodological Approach Stages

Data Collection

The dataset utilized in this study is acquired from the Central Statistics Agency (BPS) website of Central Java Province and contains poverty data for the region in 2023. It comprises 35 records and three attributes representing cities/districts and their corresponding poverty levels. The attributes include Districts and Cities in Central Java Province, which serve as the areas for clustering, the number of Uninhabitable Houses (RTLH), referring to dwellings that fail to meet safety, health, and building area spatial requirements, and the Human Development Index (HDI), which quantifies the quality of life and human development achievements within a given region. Table 1 presents the poverty dataset for Central Java Province.

Table 1. Central Java Province Poverty Dataset 2023

District/City	**Uninhabitable Houses**	**Human Development Index (%)**
Banjarnegara	14558	69.16
Banyumas	13941	73.86
Batang	11910	70.20
Blora	43212	70.63
Boyolali	5283	74.87
…	…	…
Sukoharjo	159	77.86
Tegal	1850	71.12
Temanggung	9335	71.33
Wonogiri	5449	71.97
Wonosobo	18594	70.18

Source: BPS Central Java 2023

Data Preparation

Data preparation is foundational in transforming raw and unstructured data into a clean, consistent, and analysis-ready format. This critical process encompasses several meticulous procedures, including data cleaning to remove inaccuracies, irrelevant entries, or incomplete records that may hinder analytical reliability (Fernandes et al., 2023). Additionally, normalization is applied to standardize the scale of features, ensuring comparability across variables and enhancing the accuracy of computational algorithms during analysis. By systematically organizing and refining the dataset, data preparation minimizes potential errors and establishes a robust framework that facilitates meaningful insights and supports the integrity of subsequent analytical processes. This phase is pivotal in bridging the gap between raw data and actionable knowledge, underscoring its significance in achieving reliable and impactful analytical outcomes.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14734
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Elbow Method

The Elbow technique is commonly utilized to calculate the ideal cluster count in K-means clustering. This method assesses the sum of squared errors (SSE) inside clusters, where SSE indicates the cumulative distance the distance separating data points from their corresponding cluster centers. Plotting the SSE values against the cluster count usually results in a clear 'elbow' shape at a particular graph point, signifying the ideal number of clusters. These elbow point balances minimize intra-cluster variance and avoid over-segmentation, thus ensuring an efficient clustering solution. The essential of the Elbow method lies in its ability to provide a systematic and visual means for establishing the ideal cluster count clusters and optimizing the clustering process for accuracy and interpretability (Shi et al., 2021). The formula for the Elbow method can be expressed as (1).

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \qquad (1)$$

Here:

k is the quantity of clusters.
C_i represents the data points within cluster i.
μ_i denotes the centroid of cluster i.
$\|x-\mu_i\|^2$ indicates the squared distance between each data point xx and its cluster centroid μ_i.

K-Means

Once the ideal cluster count (K) is identified utilizing the Elbow method (Liu & Deng, 2021), the subsequent step involves implementing the K-Means clustering algorithm. The K-Means algorithm operates iteratively to partition the data into K distinct clusters.

Steps of the K-Means Algorithm:

Initialization: Establish the quantity of clusters (K) using the elbow technique, then choose K centroids chosen at random from the dataset to serve as the initial cluster centers.

Cluster Assignment: Compute the distance measured by Euclidean between each data point and its centroids. Assign the data point to the cluster that has the closest centroid point.

Centroid Update: Compute the updated centroid for every grouping by computing the mean location of all data points allocated to it.

Iteration: reapply steps 2 and 3 repeatedly as far as the centroids converge, which means their positions no longer move appreciably, or when the maximum number of iterations is achieved.

This approach ensures a mathematically robust and efficient clustering process, optimizing centroids and data assignments to lower the within-cluster sum of squares (SSW) for cohesive and precise clusters.

The Euclidean distance formula provides a measure of proximity involving a data point and a cluster centroid in n-dimensional space. This involves calculating the square root of the summed squared deviations involving the data point's features and those of the centroid. This metric is fundamental in clustering algorithms like K-Means, as it ensures accurate assignment of data points to the nearest cluster by minimizing intra-cluster variance, thus contributing to robust and meaningful partitioning of datasets. The Euclidean Distance formula, as presented in (2), is defined as follows:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{n} (x_{ij} - c_{kj})^2} \qquad (2)$$

Where:

x_i: Represents the data point in consideration.
c_k: Represents the centroid of the cluster.
n: Number of dimensions or features in the data.
x_{ij}: The jth attribute of the data point x_i.
c_{kj}: The jth attribute of the cluster centroid c_k.

Model Evaluation using Silhouette Coefficient

Evaluation is a stage of model testing that assesses the accuracy of the results. This model evaluation uses the Silhouette Coefficient (3) (Lai et al., 2025). The Silhouette Coefficient span from -1 to 1. The cluster quality will improve if the average Silhouette value approaches 1. Vice versa, the cluster quality will be worse if the average Silhouette value approaches -1.

$$s = \frac{b-a}{\max(a,b)} \qquad (3)$$

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i2.14734 p-ISSN : 2541-044X

The Silhouette Coefficient and its assessment criteria comprise four evaluation levels, each with an approximate interval of 0.25. These levels include Strong Structure (0.75 < Silhouette ≤ 1.0), Medium Structure (0.5 < Silhouette ≤ 0.75), Weak Structure (0.25 < Silhouette ≤ 0.5), and No Structure (Silhouette ≤ 0.25).

RESULT

This study uses the K-Means technique to cluster poverty areas established on districts or cities in Central Java Province. The method of research includes stages such as data collection, data preparation, cluster optimization using the Elbow technique, K-Means clustering, and evaluation. The collection of poverty datasets for Central Java province is gathered from the website of the Central Statistics Agency (BPS) in 2023, which has three attributes: the district/city area of Central Java province, the Human Development Index (HDI), and Uninhabitable Houses (RTLH). The next step after data collection is to prepare the data for modelling.

Before preparing the data, descriptive statistics were used to see the distribution. We conducted a descriptive analysis to understand the characteristics of Uninhabitable Houses (RTLH) and Human Development Index (HDI) data in various regions. Table 2 provides descriptive statistics for the dataset on poverty in Central Java province.

Table 2. Statistical Descriptive

Statistics	Uninhabitable Houses	Human Development Index (%)
Count	35.000000	35.000000
Mean	10654.600000	74.810286
Std	17217.447126	4.699818
Min	6.000000	68.080000
25%	924.500000	71.390000
50%	5268.000000	73.860000
75%	12925.500000	77.185000
Max	90679.000000	84.990000

The study revealed that the distribution of RTLH exhibits a pronounced right skew, with a mean of 10.654 houses and a standard deviation of 17.217, indicating substantial variation across regions. The minimum value of RTLH was recorded at 6 houses, while the maximum value reached 90.679 houses. Meanwhile, the HDI had an average of 74.81% with a standard deviation of 4.7, indicating a more even distribution than RTLH. The HDI recorded a minimum value of 68.08% and a maximum of 84.99%. The fairly significant difference between the median and average in both variables indicates an asymmetric distribution. Moreover, the relationship between RTLH and HDI demonstrates a discernible negative correlation, suggesting that regions with a higher concentration of RTLH are inclined to exhibit lower HDI values.

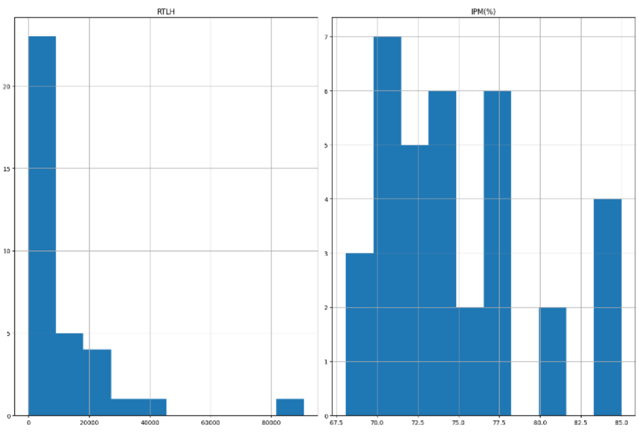

Fig 2. Histogram

We have conducted a histogram visualization to analyze the frequency distribution of the variable, providing insights into its underlying patterns and trends. Fig 2 shows the histogram of the data distribution for the HDI and HDI variables. The resulting histogram provides an overview of the distribution of Uninhabitable Houses (RTLH) and Human Development Index (HDI) data in various regions. The distribution of RTLH reveals a pronounced right-skewed pattern, characterized by most regions having relatively low RTLH counts while a few regions exhibit exceptionally high numbers. This distribution underscores a substantial disparity in housing conditions across regions.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14734
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In contrast, the distribution of HDI appears more balanced, with values consistently between 68% and 85%, suggesting a relatively even spread in human development indices. Although there are variations between regions, the HDI value tends to be in the middle range (70%—80%) without extreme outliers. This analysis indicates a negative relationship between RTLH and HDI, where regions with a high number of RTLH tend to have a lower HDI.

Table 3. Dataset After Normalization

District/City	Uninhabitable Houses	Human Development Index (%)
Banjarnegara	0.23002173	-1.21978692
Banyumas	0.19366281	-0.20514822
Batang	0.07397891	-0.99527112
Blora	1.9185606	-0.90244248
Boyolali	-0.31654064	0.03447921
…	…	…
Sukoharjo	-0.61849056	0.65837407
Tegal	-0.51884237	-0.796661
Temanggung	-0.07776212	-0.75132608
Wonogiri	-0.3067585	-0.61316251
Wonosobo	0.46785738	-0.99958874

The subsequent stage involves data preparation, wherein normalization is applied using the standard scaler method to address disparities in variable scales. The resulting dataset, post-normalization is presented in Table 3. This process ensures uniformity across variables, facilitating more accurate analysis and reliable outcomes

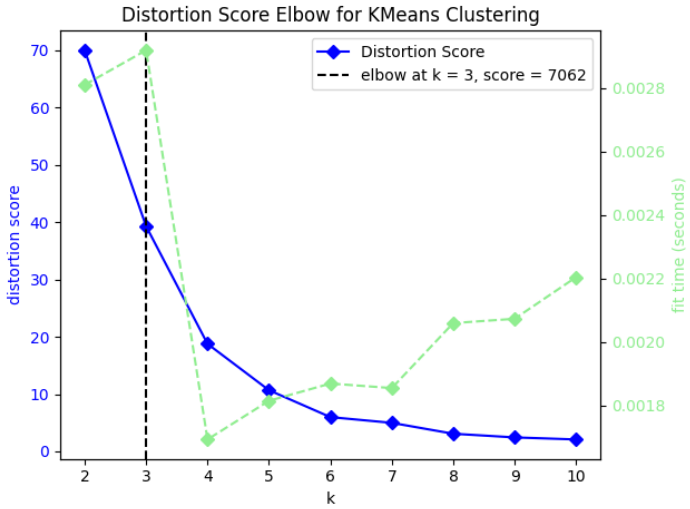

Fig 3. Elbow Method

The elbow technique establishes the quantity of clusters. Based on Fig 3, the ideal cluster count obtained is 3, with a distortion score value of 70.62. The quantity of clusters is established by the elbow point, which is the point at which the drop in the distortion score starts to declarate markedly after k = 3. This analysis shows that dividing the data into three groups is representative without losing too much information. Furthermore, the computation time (fit time) stays within the low range, ensuring the proper execution of the clustering process.

Table 4. Cluster Evaluation

Number of Cluster	Silhouette Score
2	0.458
3	0.504
4	0.479
5	0.461

The subsequent step involves evaluating the clustering outcomes utilizing the Silhouette Coefficient, a metric that assesses the standard of the cluster formation. As indicated in Table 4, cluster 3 outperforms the other clusters, achieving a silhouette score of 0.504. This score falls within the Medium Structure category, signifying that the clustering results effectively differentiate data into distinct clusters. Furthermore, the findings reflect meaningful

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14734
e-ISSN : 2541-2019
p-ISSN : 2541-044X

patterns in the relationship between the Human Development Index (HDI) and the RTLH variables, underscoring the clustering model's analytical relevance.

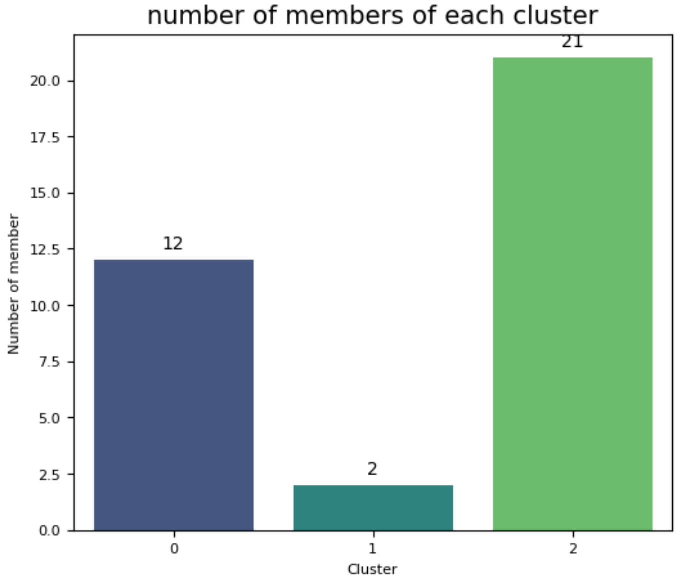

Fig 4. Frequency Distribution of Members in each Cluster

Following the evaluation of the clustering results, the subsequent step involves categorizing the members of each cluster and visualizing their distribution using a bar chart. As depicted in Fig 4, the frequency distribution highlights that cluster 0 comprises 12 regions, cluster 1 includes 2 regions, and cluster 2 encompasses 21 regions. This visualization provides a clear overview of the regional grouping within each cluster, facilitating a better understanding of the clustering outcomes.

Table 5. Cluster Regency

Cluster	Total Members	Cluster Members
0	12	Karanganyar, Klaten, Kota Magelang, Kota Pekalongan, Kota Salatiga, Kota Semarang, Kota Surakarta, Kota Tegal, Kudus, Magelang, Semarang, Sukoharjo.
1	2	Blora, Grobogan.
2	21	Banjarnegara, Banyumas, Batang, Boyolali, Brebes, Cilacap, Demak, Jepara, Kebumen, Kendal, Pati, Pekalongan, Pemalang, Purbalingga, Purworejo, Rembang, Sragen, Tegal, Temanggung, Wonogiri, Wonosobo.

Table 5 shows the provincial areas in Central Java that are included in each cluster category. Based on the clustering results, the following step is to visualize the cluster results using Principal Component Analysis (PCA) to see the distribution of areas within every cluster. Purple indicates cluster 0, green indicates cluster 1, and yellow indicates cluster 2. Cluster visualization is shown in Fig 5.

*Amiq Fahmi

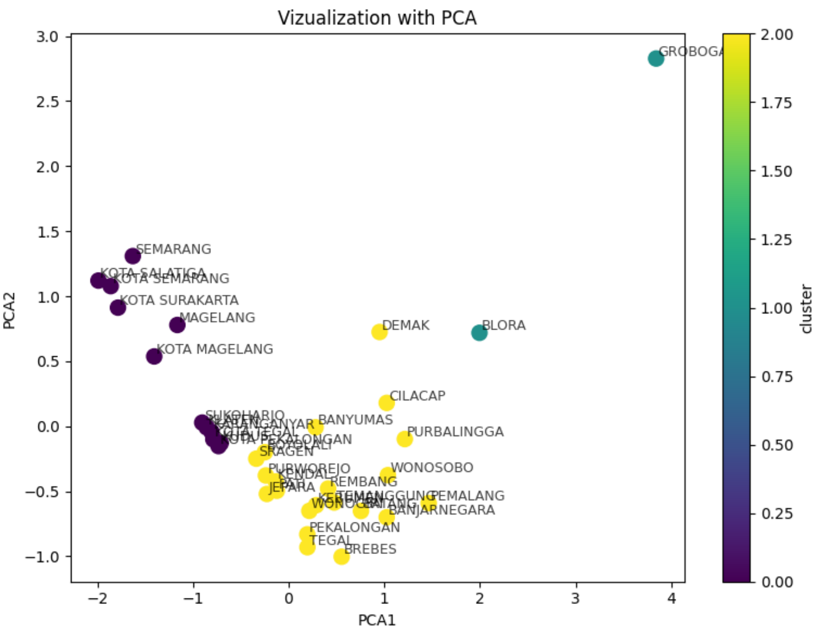

Fig 5. Visualization with PCA

Figure 5 presents the clustering results visualized using Principal Component Analysis (PCA), which reduces the dimensionality of the dataset into two principal components (PCA1 and PCA2) for easier interpretation. Evey point on the plot indicates a area or city in Central Java Province with colors indicating their respective clusters. The data is grouped into three main clusters where the first cluster (purple) is dominated by major cities such as Semarang and Surakarta, which tend to have higher Human Development Index (HDI) values. The second cluster (yellow) includes most regions with moderate characteristics, while the third cluster (greenish-blue) consists of outlier areas such as Grobogan and Blora, which may exhibit extreme values in indicators such as uninhabitable housing or HDI. This visualization provides a clearer understanding of the spatial distribution of regions based on shared characteristics.

Table 6. Descriptive Statistics for Each Cluster

Variable	Statistics	Cluster 0	Cluster 1	Cluster 2
Uninhabitable Houses	Mean	1362.333333	66945.5	10603.428571
	Std	2397.05601	33564.237583	8705.381494
	Count	12	21	21
Human Development Index (%)	Mean	80.255833	71.06	72.055714
	Std	3.401437	0.608112	2.000579
	Count	12	2	21

Table 6 is the descriptive statistics for each cluster to see the regional characteristics of each cluster. Based on Table 6, cluster 0 has the highest average Human Development Index (HDI) of 80.26% with a standard deviation of 3.40. The number of Uninhabitable Houses (RTLH) is the lowest on average, at 1.362 units, with a standard deviation of 2.397. This cluster represents a more developed area with better community welfare and minimal RTLH. Cluster 1 has the lowest average HDI of 71.06%, with a standard deviation of 0.61, indicating a reasonably uniform value in this cluster area. The number of RTLH is very high, with an average of 66.945 units and a standard deviation of 33.564, indicating a significant disparity. This cluster reflects a disadvantaged area with a low level of welfare and a considerable number of unhabitable houses. Cluster 2 The average HDI is 72.06%, with a standard deviation of 2.00, indicating that the areas in this cluster have relatively stable conditions. The number of RTLH is in the middle category, with an average of 10.603 units and a standard deviation of 8.705, which is still relatively high but not as extreme as cluster 1. This cluster represents an area with moderate development conditions, where the level of welfare still needs to be improved but not as severely as in cluster 1.

DISCUSSIONS

Utilization of the K-Means clustering algorithm to poverty data in Central Java effectively grouped districts and cities according to socio-economic variables, particularly the proportion of Uninhabitable Houses (RTLH) and the Human Development Index (HDI). The evaluation results, with a silhouette score of 0.504 at K=3, indicate a moderately well-structured clustering. These results align with previous studies such as (Khalif et al., 2024) and

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14734
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Agustina & Dana Danar, 2024), which confirmed the utility of K-Means in identifying areas with varying poverty levels and informing targeted intervention strategies.

The clustering revealed distinct socio-economic patterns: regions with high poverty levels tend to have lower HDI scores and a higher prevalence of RTLH, whereas clusters with lower poverty levels exhibit better development indicators. This classification provides valuable insights for local governments to design more effective and targeted poverty alleviation programs. Moreover, consistent use of clustering methods can support monitoring and evaluation of poverty trends over time, ensuring data-driven policy adjustments.

However, this study is limited by the scope and diversity of available data. K-Means also has limitations in handling non-linear distributions and unbalanced cluster sizes. Future research should consider integrating complementary algorithms such as DBSCAN or Hierarchical Clustering, as well as incorporating additional variables like infrastructure access and demographic factors. A more comprehensive approach could enhance the accuracy of clustering results and provide a stronger foundation for equitable and adaptive policy development.

CONCLUSION

The strategic clustering of poverty data in Central Java Province using the K-Means algorithm identifies three distinct groups: cluster 0 (12 developed districts/cities with high HDI and low RTLH), cluster 1 (2 underdeveloped districts/cities with low HDI and very high RTLH), and cluster 2 (21 intermediate districts/cities with moderate HDI and RTLH). The Elbow method effectively establishes the ideal cluster count, while the Silhouette Score evaluation (0.504 at K = 3) confirms the clustering structure within the medium category.

The findings provide strategic insights for the Central Java Provincial Government, guiding policy interventions focusing on cluster 1, where urgent welfare improvements and housing interventions are needed. By leveraging data-driven clustering, policymakers can optimize resources, improve targeting accuracy, and foster socio-economic advancements in the most vulnerable regions.

Future studies can integrate additional socio-economic indicators such as education access, healthcare availability, and infrastructure development to enhance poverty analysis. Refining clustering techniques can improve predictive accuracy and policy relevance, enabling more effective strategies for poverty alleviation.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Undergraduate Program Study in Information Systems and the guidance of academic advisors, both of which were essential to the successful publication of this research.

REFERENCES

Agustina, I., & Dana Danar, R. (2024). Implementasi K-Means Clustering untuk Pengelompokkan Tingkat Kemiskinan di Pulau Jawa Menggunakan Bahasa Pemrograman Python. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*, 945–951. https://doi.org/https://doi.org/10.36040/jati.v8i1.8858

Amelia, Nur, I. M., Rizky, M., & Milasari, S. P. (2023). Pengelompokan Tingkat Kemiskinan di Provinsi Jawa Barat dengan Metode K-Means Clustering. *Journal of Data Insights*, *1*(2), 51–61. https://doi.org/10.26714/jodi

Andriyani, F., & Puspitarani, Y. (2022). Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews. *SinkrOn*, *7*(3), 944–949. https://doi.org/10.33395/sinkron.v7i3.11569

Andriyani, W., Nasyuha, A. H., Syahra, Y., & Triaji, B. (2023). Clustering Analysis of Poverty Levels in North Sumatra Province Using the Application of Data Mining with the K-Means Algorithm. *Jurnal Media Informatika Budidarma*, *7*(4), 1971. https://doi.org/10.30865/mib.v7i4.6867

Arinto Umbu Dasa, Gergorius Kopong Pati, & Emirensiana Dappa Ege. (2024). Penerapan Algoritma K-Means Clustering Data Penduduk Miskin Berdasarkan Desa di Kecamatan Tana Righu. *Polygon : Jurnal Ilmu Komputer Dan Ilmu Pengetahuan Alam*, *2*(6), 1–7. https://doi.org/10.62383/polygon.v2i6.245

Azzahro, Z., Chyntia, L. F. A. R., Nasar, M., & Widianto, W. (2024). Implementation of the K-Means Clustering Method for Grouping Poverty Data: Case Studies in Indonesia. *AIP Conference Proceedings*, *2927*(1). https://doi.org/https://doi.org/10.1063/5.0193516

Bahauddin, A., Fatmawati, A., & Sari Permata, F. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *MISI (Jurnal Manajemen Informatika & Sistem Informasi)*, 2614–3739.

Dubey, P., & Rajavat, A. (2023). Effective K-Means Clustering Algorithm for Efficient Data Mining. *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, 1–6. https://doi.org/https://doi.org/10.1109/ViTECoN58111.2023.10157179

Fernandes, A. A. A., Koehler, M., Konstantinou, N., Pankin, P., Paton, N. W., & Sakellariou, R. (2023). Data Preparation: A Technological Perspective and Review. *SN Computer Science*, *4*(4). https://doi.org/https://doi.org/10.1007/s42979-023-01828-8

*Amiq Fahmi

Khalif, A., Nur Hasanah, A., Hafizh Ridwan, M., & Nurina Sari, B. (2024). Klasterisasi Tingkat Kemiskinan di Indonesia Menggunakan Algoritma K-Means. In *Generation Journal* (Vol. 8, Issue 1). https://ojs.unpkediri.ac.id/index.php/gj/article/view/21470/3747

Lai, H., Huang, T., Lu, B., Zhang, S., & Xiaog, R. (2025). Silhouette Coefficient-Based Weighting K-Means Algorithm. *Neural Computing and Application*, *37*(5), 3061–3075. https://doi.org/https://doi.org/10.1007/s00521-024-10706-0

Lasfeto, D. B., Setyorini, T., Mauta, J. J., Bria, M., & Nenobais, O. O. N. (2024). A Simple Classification and Clustering of Poverty in Rural Areas Using Machine Learning. *Journal of Infrastructure, Policy and Development*, *8*(8), 5938. https://doi.org/10.24294/jipd.v8i8.5938

Liu, F., & Deng, Y. (2021). Determine the Number of Unknown Targets in Open World Based on Elbow Method. *IEE Transactions on Fuzzy Systems*, *29*(5), 986–995. https://doi.org/https://doi.org/10.1109/TFUZZ.2020.2966182

Lubis, S. S., & Hendrik, B. (2023). Implementasi Data Mining Pengelompokan Data Penjualan Berdasarkan Pembelian dengan Menggunakan Algoritma K-Means Pada UD.Martua. *Journal of Information System and Education Development*, *1*(3), 36–41.

Maesarah, S., & Fatah, Z. (2024). Application of K-Means Clustering for Grouping Regions Based on Poverty Levels in Indonesia. *Multidisciplinary Scientific Journal*, *6*, 114–120. https://doi.org/10.69714/3d5mkb02

Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset. *Applied Sciences (Switzerland)*, *12*(15). https://doi.org/10.3390/app12157515

Purwanti, E. (2024). Analisis Deskriptif Profil Kemiskinan Indonesia Berdasarkan Data BPS Tahun 2023. *Jurnal Mahasiswa Humanis*, *4*(1), 1–10.

Qulsum, U., Anggraini, J., & Putri, E. (2024). Analysis of Poverty as A Main Problem of the Indonesian Economy and Its Overcome. *Journal of Social and Library Science*, *2*, 60–66.

R, N. N. F., Anggraeni, D. S., & Enri, U. (2022). Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient. *Jurnal Teknologi Informasi Komunsikasi*, *9*, 29–35. https://doi.org/10.38204/tematik.v9i1.921

Riyono, J., & Pujiastuti, C. E. (2022). Simulation Of The K-Means Clustering Algorithm with The Elbow Method in Making Clusters of Provincial Poverty Levels in Indonesia. *Jurnal Matematika MANTIK*, *8*(2), 113–123. https://doi.org/10.15642/mantik.2022.8.2.113-123

Sepriyanti, N., Sani Nahampun, R., Zikri, M. H., Ambarani, I., & Rahmadeyan, A. (2022). Implementation of K-Means Clustering to Group Poverty Levels in Riau Province. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 59–65. https://journal.irpi.or.id/index.php/sentimas

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A Quantitative Discriminant Method of Elbow Point for The Optimal Number of Clusters in Clustering Algorithm. *Eurasip Journal on Wireless Communications and Networking*, *2021*(1). https://doi.org/10.1186/s13638-021-01910-w

Vanessa, K., Iswanto, I. A., Wijaya, K., & Hidayat, M. F. (2024). Comparing K-Means and DBSCAN Algorithm for Clustering Poverty Levels in Papua Islands. *2024 9th International Conference on Information Technology and Digital Applications (ICITDA)*, 1–6. https://doi.org/https://doi.org/10.1109/ICITDA64560.2024.10810077

Wirayudha, A. B., & Prasetyia, F. (2024). Analisis Determinan Kemiskinan di 34 Provinsi di Indonesia. *Journal of Development Economic and Social Studies*, *3*(3), 763–779. https://doi.org/10.21776/jdess.2024.03.3.08